# Safe Adaptation in an Automotive Vehicle: The Driver Advocate™

**Bradford W. Miller, Noel Massey, Robert M. Gardner**

Motorola, Inc.
{Bradford.W.Miller, Noel.Massey, MikeGardner}@motorola.com

## Our Domain

We are investigating the issues surrounding information presentation to a vehicle operator, to enhance their environmental awareness and provide better on-task performance, despite the presence of numerous distractions. Two trends have caused us to focus on this area. First, the increased availability of inexpensive sensors such as GPS, radar and video that can provide additional services to a driver, such as navigation assistance, automatic cruise control, and collision warnings. Second, drivers are tending to spend more time in their vehicle, and wish to make productive use of this time, either in office productivity tasks, or maintaining external relationships (e.g., using a cellular phone to participate in meetings or engage in social conversations).

As the number and complexity of in-vehicle (or driver portable, e.g., bringing a portable CD player into the vehicle) services increases, the potential for Fatal Distraction™ also increases. While we have chosen this domain for the wide variety of experiments in human factors, interface design, and service architectures it affords, our current particular focus of attention is on an intelligent system that will monitor driver attention, and aid in appropriate information presentation to reduce distraction, and improve driver situation awareness.

## Our Problem

We have chosen to use an intelligent agent architecture for a number of the beneficial attributes it affords. In particular, given our focus on driver assistance, we are attracted to the potential for robustness to unanticipated input, dynamic reconfiguration and resource optimization, as well as the ability to distribute both computational resources and human software engineering to rapidly prototype systems which can be used for human interface experiments (see, e.g., [Weiss 1999]). Because every individual driver has different capabilities and every vehicle performs differently, it is important for the system to adapt to the particular driver/vehicle combination (e.g., including sensor characteristics) it is presented with. With time, driver performance and vehicle characteristics can be expected to change as well, so it is important that we continue to adapt, rather than simply register a given level of performance for long-term use.

We have identified a number of areas for adaptation using machine learning techniques:

- the operating model
- driver task preferences
- driver capability
- short term behavior.

Let us explain each of these in turn. By the "operating model", we mean that as system characteristics change (perhaps because sensors are replaced), we need to update our model of how the system itself operates. We can think about short term adaptation, e.g., dealing with a sensor that currently doesn't work, and long term adaptation, e.g., dealing with a sensor that has been replaced by a different component during routine maintenance.

By "driver task preferences," we mean the preferences a driver has to achieve particular goals, including their preferred interaction with the system. For instance, some drivers may prefer visual alerts using a heads-up display, while others may prefer audible alerts in most cases. Or some drivers may not wish to be informed of minor lane deviation, finding it distracting.

"Driver capability" has to do with our model of how well the driver will perform in a given situation. We need to learn this from the driver's behavior in similar situations. For instance, some drivers may be able to execute a particularly sharp turn at a relatively high speed, while others would put the car into a skid, or begin to steer into the curve too soon.

Adapting to "short term behavior" has to do with how the systems' normal behavior has to change regardless of other preferences because of a temporary unusual situation. For instance, while the driver might normally be notified that they are approaching a car too quickly by enhancing the image of the brake lights on the vehicle, an audible alert might need to be used if the driver does not currently have their attention on the road ahead (perhaps distracted by the buttons on their radio).

While our system will attempt to aid the driver in focusing on the driver task by keeping them apprised of the salient aspects of their current situation, we do not seek to construct an autonomous vehicle. The driver will always be firmly in charge of making the ultimate decision as to what should be done in a given situation, and control of the vehicle will remain with the driver. Our difficulty arises in making sure that any adaptation still results in a system that is "safe," that is, does not recommend choices that would be more likely to lead to the driver making an incorrect choice than had the system not been present. That is, we want the system to reduce the effects of distraction caused by the various services in the vehicle by bringing up the driver short when they appear to not be paying attention to a critical situation, and not be part of the general cacophony in the vehicle as well.

## Our Strategy

Our current strategy for approaching the problem is a combination of two tactics. First, we will have a formal specification of our agents, e.g., similar to that described in [Wooldridge 2000]. We will assume that changes to our agents proposed by those agents charged with adaptation using machine learning techniques will use rule induction (e.g., [Mitchell 1997] [Tan 1993]), and that these induced rules will also be represented in this formal language. That is, our adapted agent will be the result of taking a formal specification of an agent, adding or changing rules within it, and as a result producing a new specification of an agent. This specification can then be "compiled" or simply interpreted using a production system, if one ignores the computational complexity, and hence non real-time nature of such an approach.

For any agent specification, part of the specification will be distinguished, and consist of constraints on the behavior of the agent. To the extent we can prove these constraints correct, well and good, and constraints that cannot be proven will be added to a dynamic operating model, which we will use to monitor the agent's performance.

For a given interaction with a service, we will have the choice of enlisting an adapted agent, or an agent from which the adaptation was based, presumably, known with some degree of assurance to operate correctly in a given situation, if not optimally. Until we have sufficient experience with the new agent, we will enlist the services of both agents, just in case the new agent does not respond in time, or violates some other constraint expressed in the operating model. (The quality of the solution is presumably not something which we can easily check *a priori*, since it will require analyzing the behavior of the system and driver *post hoc*). Once we have sufficient confidence that the new version of a particular agent is operating properly, we can drop the need to recruit both versions for a service invocation.

## Open Questions

At this point, we are beginning a new project, and a number of questions come up, such as:

- When do we decide we have enough experience with a given agent to trust its future outputs (that is, no longer attempt to recruit multiple providers for the same service)?
- How do we decide which characteristics of a particular agent to measure on a given occasion, given limited computational resources?
- How do we decide which constraints directly influence safety (e.g., of the driver or of surrounding traffic and pedestrians)? Can we (automatically) learn these relationships and induce new ones?

## References


Mitchell, Tom, 1997. *Machine Learning*, Boston, Mass.: McGraw-Hill. ISBN 0-07-042807-7

Tan, Ming, 1993. Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents. In Proceedings of the Tenth International Conference on Machine Learning, Amherst, MA. 330—337.

Weiss, Gerhard, ed. 1999. *Multiagent Systems.* Cambridge, Mass.: The MIT Press. ISBN 0-262-23203-0

Wooldridge, Michael, 2000. *Reasoning About Rational Agents.* Cambridge, Mass.: The MIT Press. ISBN 0-262-23213-8